## REMARKS

The Examiner has withdrawn Claim 9 as an independent and distinct invention.

Claim 8 is allowed. Applicants appreciate the favorable examination of claim 8.

Anticipation rejection of Claim 1

Claim 1 and 3 are rejected as anticipated by Zhou, US Patent 7,222,074. Claim 3 is cancelled rendering the rejection moot.

The Examiner notes that the system of Zhou selects an acoustic model from a plurality of acoustic models (see Figure 2 and "acoustic model selection mechanism 240"). The applicants respectfully submit that the Examiner errs in considering that such selection is made from instructions contained within a voice command application, as claimed in claim 1.

Rather, in Zhou, the selection of one of the acoustic models 220A, 220B and 220C is made by an acoustic model selection mechanism on the basis of real-time input in the form of a determination of the psycho-physical state of the user, via a psycho-physical detection mechanism 110, as indicated in Figure 2 at 115. The selection is not made by instructions *within the voice command application*, e.g., by a block of code such as VXML metadata in the voice command application.

The Zhou reference explains the acoustic model selection as follows:

> In FIG. 1, a psycho-physical state sensitive spoken dialogue system 100 comprises a psycho-physical state detection mechanism 110 and a voice responding dialogue mechanism 120. The psycho-physical state detection mechanism 110 takes the speech of a user as input speech data 105 and detects the psycho-physical state 115 of the user from the input speech data 105. Such detected psycho-physical state 115 is used by the voice responding dialogue mechanism 120 to generate a psycho-physical state sensitive voice response 125. . . . (Col. 3 lines 16-25.)
>
> The psycho-physical state 115 may include mental stress or physical stress. For example, anger may be considered as mental stress and cold may be considered as physical stress. The psycho-physical state of a user may affect the acoustic properties of the user's speech. For example, when a user is

mentally stressed (e.g., angry), the loudness or the speed of the speech may increase. Acoustically, such increase may be correlated with the rise of the pitch of the voice. . . . (Col. 3 lines 40-47.)

There are many ways to detect the psycho-physical state of a person. For example, anger may be detected from a person's facial expression or physical gesture. In the present invention, the psycho-physical state detection mechanism 110 detects the psycho-physical state 115 from the voice of a person (e.g., from the input speech data 105). Such detection may be based on the acoustic characterizations of a person's voice under different psycho-physical states. . . . (Col. 4 lines 4-12.)

The speech understanding mechanism 210 includes a plurality of acoustic models 220 comprising acoustic model 1, 220a, acoustic model 2, 220b, . . . , acoustic model i, 220c . . . . ***An acoustic model selection mechanism 240 selects, based on the detected psycho-physical state of the user, appropriate acoustic models to be used in recognizing the spoken words from the input speech data 105.*** . . . (Col. 4 lines 53-59.)

FIG. 3 is an exemplary flowchart of a process, in which the psycho-physical state sensitive spoken dialogue system 100 carries out a dialogue with a user based on the psycho-physical state of the user. The input speech data 105 is received first at act 310. Based on the input speech data 105, the psycho-physical state detection mechanism 110 detects, at act 320, the current psycho-physical state of the user. For example, the detection mechanism 110 may determine that the user is frustrated (may due to some misunderstood dialogue) and such a decision may be concluded according to the acoustic characteristics of the input speech data 105.

***Based on the detected psycho-physical state 115, the acoustic model selection mechanism 240 selects, at act 330, one or more acoustic models that characterize the acoustic properties correlated with the detected psycho-physical state.*** (Col. 5 lines 41-56.)

Thus, in Zhou, the selection of the acoustic model is not made by instructions ***in the voice command application itself,*** as claimed in claim 1, but rather is made in response to input from the psycho-state detection mechanism. A voice command application in Zhou would not be able to contain instructions to select the acoustic model since the application in Zhou would not know in advance what a given user's psycho-physical state would be in advance. As the input for selection of the acoustic model in Zhou is ***clearly coming from a source outside of the voice command application.*** Zhou does not anticipate claim 1.

Obviousness rejection of claim 2

Claim 2 stands rejected as obvious over Zhou in view of Thomas et al., US 7,171,361.

The Examiner contends that Thomas' teaching of VMXL metadata tags indicting location of grammars, when combined with Zhou, would render claim 2 obvious. Claim 2 depends from claim 1 and adds that the instructions which select the acoustic model are in the form of a VXML metadata element.

While the Thomas teaching is relevant to how an application may obtain grammars (allowed utterances at a given navigation point in an application), Thomas does not suggest that the VXML tags should invoke a specific acoustic model. Grammars and acoustic models are entirely different things. The "grammar" that is being referred to in Thomas is a set of words or groups of words which are accepted as valid responses at particular navigation points in the application. Words that are not in this set of grammar are considered "out of grammar" responses, a concept well known in the art. This is what is meant by the term "grammar" in the speech recognition art. See specification, e.g., page 3 line 19 to page 4 line 2. See also the documents attached to the applicants' previous response: Nuance Speech Recognition System Version 7.0 Grammar Developer's Guide, chapter 1 pages 10-11 (pointing out the difference between the acoustic model and the grammar set in a speech recognition system), Chapter 2, page 13 (giving examples of grammar "Yes' and "No"); Chapter 3 pages 33-34 (describing choosing an acoustic model set to use with a grammar, clearly indicating that the two concepts are distinct). Speech Recognition Grammar Specification for the W3C Speech Interface Framework, section 1.1 Grammar Processor ("a speech recognizer is a grammar processor with the following inputs and outputs: * Input: A grammar or multiple grammars as defined by this specification. These grammars inform the

recognizer of the words and patterns of words to listen for. . . ."); Technology Reports W3C Speech Recognition Grammar Specification ([The W3C Speech Recognition Grammar Markup Language Specification] "defines the syntax of grammar representation. The grammars are intended for use by speech recognizers and other grammar processors so that developers can specify the words and patterns of words to be listened for by a speech recognizer.") Wikipedia definition of Acoustic Model: Background ("Speech recognition engines require two types of files to recognize speech. They require an acoustic model, which is created by taking audio recordings of speech and their transcriptions (taken from a speech corpus), and 'compiling' them into a statistical representations of the sounds that make up each word (through a process called 'training'). They also require a language model or grammar file. A language model is a file containing the probabilities of sequences of words. A grammar file is a much smaller file containing sets of predefined combinations of words.")

Zhou teaches that the acoustic model is selected on the basis of real-time input as to the user's current psycho-physical state. Combining Zhou with Thomas would suggest at most that the system of Zhou would continue to use Zhou's method of selection of an acoustic model from input as to the current psycho-physical state of the user and reference to idiom grammars via VXML metadata tags as in Thomas. That combination does not suggest including code in the application itself which selects a particular acoustic model to use. Accordingly, the two references do not render claim 2 obvious.

Obviousness rejection of claim 5

Claims 5 is rejected as obvious over Zhou in view of Kuroiwa.

Claim 5 depends from claim 1 and recites that the voice command application includes instructions which select the acoustic model based on an area code and/or local exchange number of a user accessing the application. The claim is based on the understanding that geographic location of where the user is located can be at least loosely be correlated to the type of speech or accent they may have (e.g., Southern, Maine) and an acoustic model can be selected from such location area obtained from area code or local exchange number.

Zhou, as explained above, does not involve an application itself containing instructions which select an acoustic model. Rather, acoustic model selection is made from information coming outside of the application, and in particular input from a module detecting the current psycho-physical state of the user. As explained at length in applicant's previous response, Kuroiwa deals with the selection of acoustic model on a system-wide basis though the use of a line data analyzer and a switching element. Kuroiwa (and Zhou) does not leave the selection of an acoustic model to the application developer, or code the selection of an acoustic model in the voice command application itself, but rather handles it on a system level. Thus, neither Kuroiwa nor Zhou (nor Thomas for that matter) singly or in combination teach or suggest the broader invention of claim 1, or the specific application claimed in claim 5 in which the voice command application itself includes instructions which selects the acoustic model based on area code or local exchange number of the user.

Favorable reconsideration of the application is requested.

Respectfully submitted.

McDonnell Boehnen Hulbert & Berghoff LLP

Date: 12/18/07 By: _____

Thomas A. Fairhall
Reg. No. 34591